

С.И. Носков¹, В.Н. Городенко¹

¹*Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

ПРОГРАММНЫЙ КОМПЛЕКС АНТИРОБАСТНОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ ЛИНЕЙНОГО РЕГРЕССИОННОГО УРАВНЕНИЯ

Аннотация. В статье рассматривается способ оценивания параметров линейных регрессий методом антиробастного оценивания, основанного на минимизации расстояния Чебышева между расчетными и фактическими значениями зависимой переменной на предыстории процесса и сводящегося к задаче линейного программирования. Приводится описание разработанного программного комплекса для автоматизированной оценки параметров на конкретном примере.

Ключевые слова: линейная регрессия, метод антиробастного оценивания, расстояние Чебышева, задача частично-булевого линейного программирования.

S.I. Noskov¹, V.N. Gorodenko¹

¹*Irkutsk State Transport University, Irkutsk, Russia*

SOFTWARE COMPLEX OF ANTIROBASTIC EVALUATION OF PARAMETERS OF A LINEAR REGRESSION EQUATION

Abstract. The article discusses a method for estimating linear regression parameters by the anti-bass estimation method, based on minimizing the Chebyshev distance between the calculated and actual values of the dependent variable on the process history and reduced to the linear programming problem. A description of the developed software package for automated parameter estimation using a specific example is given.

Keywords: linear regression, anti-robust estimation method, linear programming problem, scipy module.

Введение. Рассмотрим линейное регрессионное уравнение

$$y_k = \sum_{i=1}^m \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, n}, \quad (1)$$

где y – зависимая, а x_i – i -ая независимая переменные; α_i – i -ый подлежащий оцениванию параметр; ε – ошибки аппроксимации, k – номер наблюдения, n – число наблюдений (длина выборки).

Представим уравнение (1) в матричной форме:

$$y = X\alpha + \varepsilon, \quad (2)$$

где $y = (y_1, \dots, y_n)^T$, $\alpha = (\alpha_1, \dots, \alpha_m)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$, X – $(n \times m)$ – матрица с компонентами x_{ki} .

Присутствие в уравнениях (1), (2) аппроксимационной составляющей может быть вызвано:

- неточностями в фиксации значений эндогенной и экзогенных переменных;
- игнорированием каких-либо существенных факторов;
- влиянием неучитываемых помех;
- неудачным выбором формы связи между переменными.

Статья основана на работах [1-18].

Широкий класс методов оценивания параметров уравнения (2) связан с расчетом так называемых L_ν – оценок, вычисляемых посредством минимизации функций потерь вида [1]:

$$J_\nu(\alpha) = \sum_{k=1}^n |\varepsilon_k|^\nu. \quad (3)$$

Каждая из этих оценок характеризуется различной реакцией на так называемые выбросы – наблюдения, несогласующиеся со всей выборкой в целом. При этом чем больше значе-

ние ν , тем сильнее L_ν – оценка реагирует на выбросы. В регрессионном анализе методы оценивания, слабо реагирующие на выбросы, или вообще их игнорирующие, называют робастными.

Методом оценивания параметров уравнения (2), соответствующим $\nu \rightarrow \infty$, является метод антиробастного оценивания (МАО), соответствующий расстоянию Чебышева. Отметим, что в соответствии с хорошо известным фактом имеет место соотношение:

$$J_\infty = \max_{k=1,n} |\varepsilon_k|. \quad (4)$$

В [1] показано, что МАО – оценка может быть определена с помощью решения следующей задачи линейного программирования (ЛП).

$$X\alpha^1 - X\alpha^2 + u - \nu = y, \quad (5)$$

$$u_k + \nu_k - r \leq 0, \quad k = \overline{1, n}, \quad (6)$$

$$\alpha^1 \geq 0, \quad \alpha^2 \geq 0, \quad u \geq 0, \quad \nu \geq 0, \quad r \geq 0, \quad (7)$$

$$r \rightarrow \min. \quad (8)$$

МАО – оценка линейной регрессии (2) рассчитывается по формуле

$$\alpha = \alpha^1 - \alpha^2$$

после решения задачи ЛП (5) – (8) с $2(n+m)+1$ переменными и $2n$ ограничениями.

Описание программного комплекса (ПК МАО) антиробастного оценивания параметров линейного уравнения регрессии

Для автоматизированной оценки параметров регрессионных моделей был разработан программный комплекс на языке программирования Python версии 3.

Создание графического интерфейса производилось при помощи подключения модуля Tkinter. Для удобной работы с матрицами использовалась библиотека numpy, а для автоматизированного решения задачи ЛП подключалась библиотека scipy.

Общий процесс работы программного комплекса можно описать следующей последовательностью шагов.

1. Ввод исходных данных.
2. Настройка программы.
3. Формирование задачи линейного программирования.
4. Решение задачи линейного программирования.
5. Вывод полученных результатов.

Ввод исходных данных в программу осуществляется путем загрузки двух текстовых файлов с расширением *.txt, соответствующих матрице X и вектору y . Первый файл должен содержать столбцы значений независимых переменных, второй – столбец значений зависимой переменной. Столбцы разделяются между собой пробелом, а строки - при помощи знака переноса строки. Примеры таких файлов представлены на рис. 1. и рис. 2 соответственно.

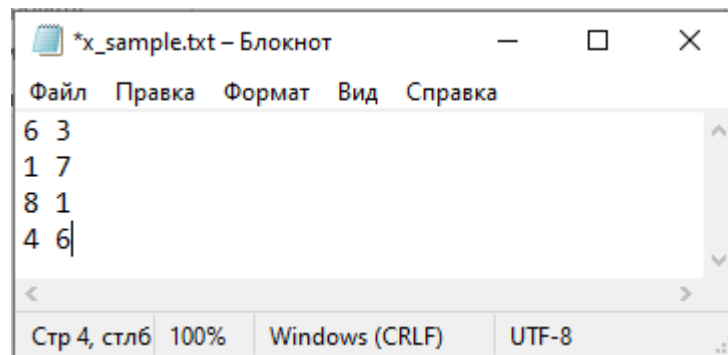


Рис. 1. Пример файла с независимыми переменными

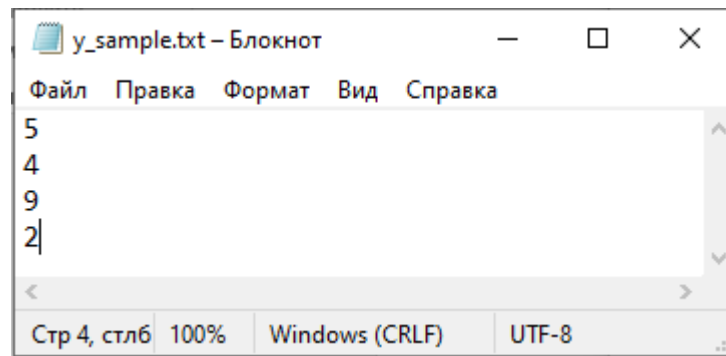


Рис. 2. Пример файла с зависимой переменной

Помимо загрузки из файла программный комплекс предусматривает возможность ручного ввода данных в предназначенных для этого объектах текстовых полей.

Настройка программного комплекса производится в отдельном окне. Такое окно показано на рис. 3. В нём разрешается настраивать некоторые параметры работы программы. Первый параметр – точность. Регулировка точности позволяет предусмотреть количество знаков после точки для вещественных чисел. По умолчанию выводится 4 знака, изменять их количество можно в диапазоне от 1 до 10. Помимо точности можно менять так называемый допуск решения - от 0.01 до 1e-10. Кроме того, можно выбрать предпочитаемый пользователем метод решения задачи ЛП - симплекс-метод или метод внутренней точки.

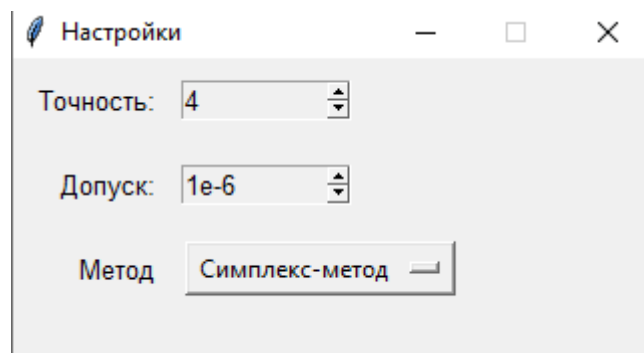


Рис. 3. Окно настроек программного комплекса

Решение задачи линейного программирования производится при помощи библиотеки `scipy` и её модуля `linalg`. Указанная библиотека не только является наиболее популярной в использовании, но и имеет хорошую документацию и описание всех заложенных функций и процедур.

Решение задачи ЛП в программном комплексе реализует функция `linprog` модуля `linalg`. Функция `linprog` принимает следующие параметры:

- вектор целевой функции c ;
- матрицу наблюдений независимых переменных A_{ub} , для которых задаются ограничения – неравенства «меньше или равно»;
- вектор наблюдений зависимой переменной b_{ub} , который соответствует матрице A_{ub} ;
- матрицу наблюдений независимых переменных A_{eq} , для которых задаются ограничения – равенства;
- вектор наблюдений зависимой переменной b_{eq} , который соответствует матрице A_{eq} ;
- метод решения;
- дополнительные опции (количество итераций, допуск и прочие).

Формирование и решение задачи ЛП инициализируется путем нажатия кнопки «Найти решение» на главном окне программного комплекса.

В случае успешного решения в главном окне появится результат, в противном случае - сообщение об его отсутствии. Результат решения включает в себя вектор искомых параметров α , вектор ошибок аппроксимации ε , а также значения следующих критериев адекватности уравнения:

- сумма модулей ошибок (S);
- средняя относительная ошибка аппроксимации (E);
- максимальная по модулю ошибка (r);
- количество максимальных ошибок (H).

Помимо вывода на экран все составляющие результата решения при необходимости можно сохранить в файл в формате *.txt.

Процесс решения задачи в ПК MAO

Работу программного комплекса имеет смысл рассмотреть на конкретном примере с данными, представленными на рис. 1-3. Для удобства восприятия данные приведены также на рис. 4.

При помощи кнопок «Импорт из файла» на главном окне занесем данные в соответствующие текстовые поля. Установим галку «св. член». Сохраним окно настроек по умолчанию. Программный комплекс готов к расчету.

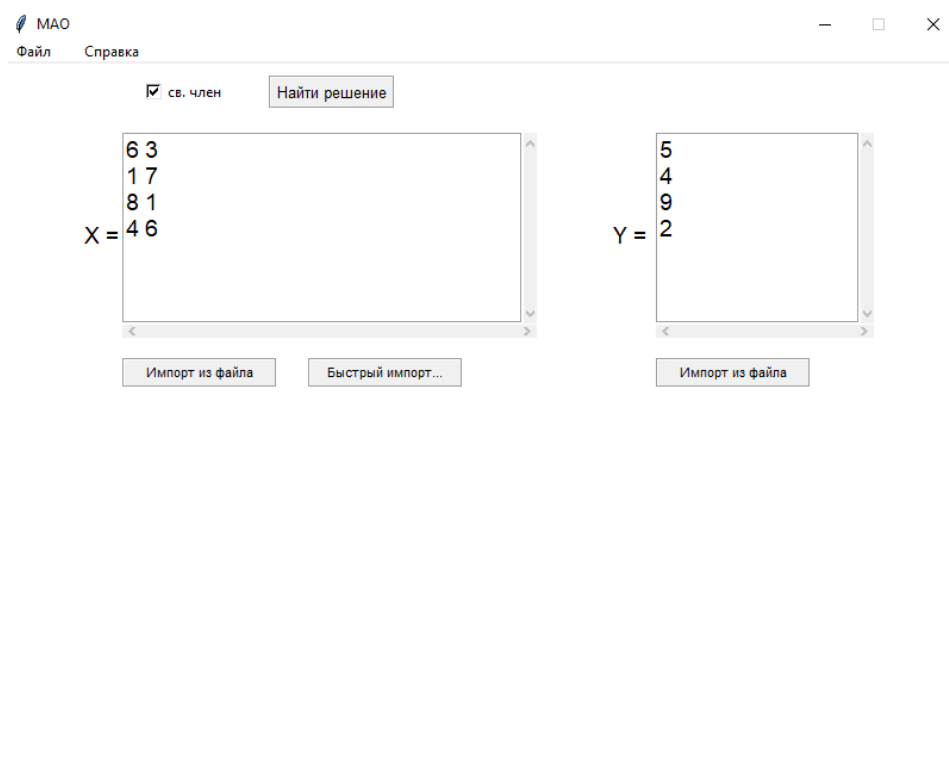


Рис. 4. Главное окно программного комплекса

Нажмем кнопку «Найти решение». В нижней половине главного окна отобразятся все необходимые параметры. Результат работы программы показан на рис. 5.

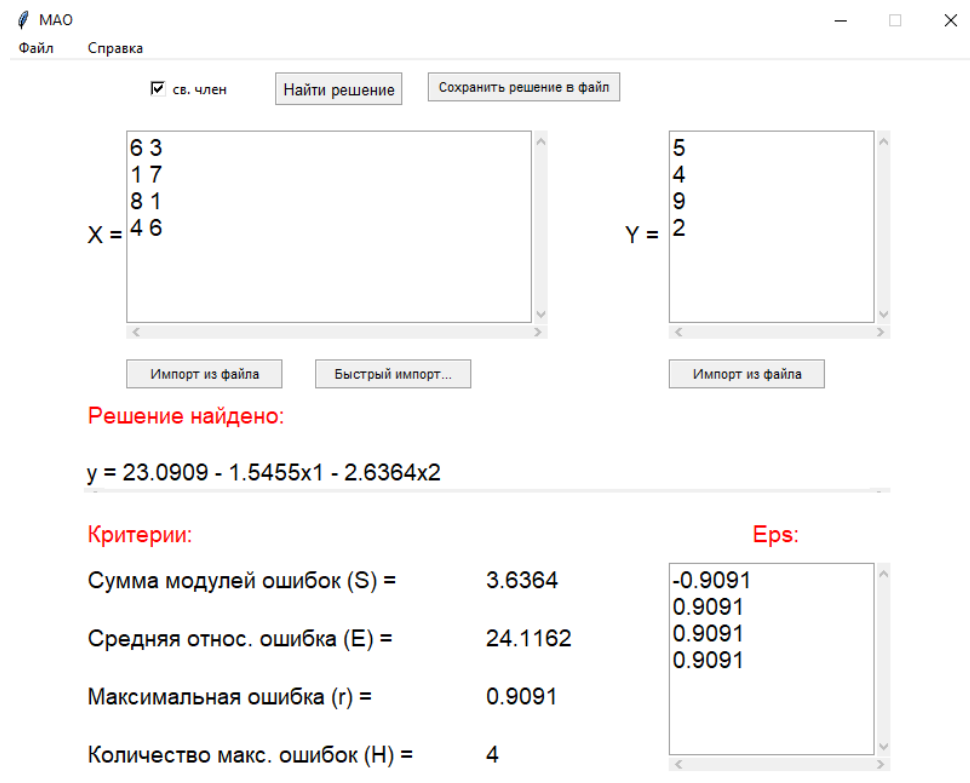


Рис. 5. Результат работы программного комплекса MAO

Заключение. В дальнейшем авторы намерены развивать программный комплекс. Планируется, в частности, расширение его возможностей по вводу данных с использованием форматов *.xlsx, *.docx, *.csv, а также увеличение числа рассчитываемых значений критериев адекватности регрессионного уравнения (1).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. – Иркутск: Облформпечать. – 1996. – 320 с.
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Юнити, 1998. – 1022 с.
3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Книга 1. В 2-х кн. – М.: Финансы и статистика, 1986. – 366 с.
4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Книга 2. В 2-х кн. – М.: Финансы и статистика, 1986. – 351 с.
5. Себер Дж. Линейный регрессионный анализ. – М.: Издательство «Мир», 1980. – 456 с.
6. Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 2009. – 465 с.
7. Носков С.И., Лоншаков Р.В. Идентификация параметров кусочно-линейной регрессии//Информационные технологии и проблемы математического моделирования сложных систем. – 2008. – № 6. – С. 63-64.
8. Ильина Н.К., Лебедева С.А., Носков С.И. Идентификация параметров некоторых негладких регрессий//Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – № 17. – С. 111.
9. Носков С.И. Критерий «согласованность поведения» в регрессионном анализе//Современные технологии. Системный анализ. Моделирование. – 2013. – №1(37). – С.107-110.
10. Носков С.И. Построение эконометрических зависимостей с учетом критерия «согласованность поведения»//Кибернетика и системный анализ. – 1994. – № 1. – С. 177.

11. Kreinovich V., Lakeyev A.V., Noskov S.I. Approximate linear algebra is intractable // *Linear Algebra and its Applications*. – 1996. – Т. 232. – № 1-3. – С. 45-54.
12. Базилевский М.П., Носков С.И. Алгоритм формирования множества регрессионных моделей с помощью преобразования зависимой переменной // *Международный журнал прикладных и фундаментальных исследований*. – 2011. – № 3. – С. 159-160.
13. Носков С.И. Точечная характеристика множества Парето в линейной многокритериальной задаче // *Современные технологии. Системный анализ. Моделирование*. – 2008. – № 1 (17). – С. 99-101.
14. Lakeyev A.V., Noskov S.I. A description of the set of solutions of a linear equation with interval defined operator and right-hand side // *Doklady Mathematics*. – 1993. – Т. 47. – № 3. – С. 518.
15. Лакеев А.В., Носков С.И. Метод наименьших модулей для линейной регрессии: число нулевых ошибок аппроксимации // *Современные технологии. Системный анализ. Моделирование*. – 2012. – № 2 (34). – С. 48-50.
16. Носков С.И. Обобщенный критерий согласованности поведения в регрессионном анализе // *Информационные технологии и математическое моделирование в управлении сложными системами*. – 2018. – № 1 (1). – С. 14-20.
17. Носков С.И. О методе смешанного оценивания параметров линейной регрессии // *Информационные технологии и математическое моделирование в управлении сложными системами*. – 2019. – № 1 (2). – С. 41-45.
18. Носков С.И. Идентификация параметров кусочно-линейной функции риска // *Транспортная инфраструктура Сибирского региона*. – 2017. – Т. 1. – С. 417-421.

REFERENCES

1. Noskov S.I. Technology for modeling objects with unstable functioning and uncertainty in data. – Irkutsk: Oblinformpechat. – 1996. – 320 p.
2. Ayvazyan S.A., Mkhitaryan V.S. Applied statistics and fundamentals of econometrics. – M.: Unity, 1998. – 1022 p.
3. Draper N., Smith G. Applied regression analysis. Book 1. In 2 book. – M.: Finance and Statistics, 1986. – 366 p.
4. Draper N., Smith G. Applied regression analysis. Book 2. In 2 book. – M.: Finance and Statistics, 1986. – 351 p.
4. Seber J. Linear regression analysis. – M.: Mir Publishing House, 1980. – 456 p.
5. Dougherty K. Introduction to Econometrics. – M.: INFRA-M, 2009. – 465 p.
6. Noskov S.I., Lonshakov R.V. Identification of parameters of piecewise linear regression // *Information technologies and problems of mathematical modeling of complex systems*. – 2008. – No. 6. – Pp. 63-64.
8. Ilyina N.K., Lebedeva S.A., Noskov S.I. Identification of parameters of some non-smooth regressions // *Information technologies and problems of mathematical modeling of complex systems*. – 2016. – No. 17. – P. 111.
9. Noskov S.I. The criterion of “consistency of behavior” in the regression analysis // *Modern technologies. System analysis. Modeling*. – 2013. – No. 1 (37). – Pp. 107-110.
10. Noskov S.I. Construction of econometric dependencies taking into account the criterion of “consistency of behavior” // *Cybernetics and system analysis*. – 1994. – No. 1. – P. 177.
11. Kreinovich V., Lakeyev A.V., Noskov S.I. Approximate linear algebra is intractable // *Linear Algebra and its Applications*. – 1996. – Т. 232. – No. 1-3. – Pp. 45-54.
12. Bazilevsky M.P., Noskov S.I. Algorithm for the formation of a multitude of regression models by transforming a dependent variable // *International Journal of Applied and Fundamental Research*. – 2011. – No. 3. – Pp. 159-160.
13. Noskov S.I. Point characterization of the Pareto set in a linear multicriteria problem // *Modern Technologies. System analysis. Modeling*. – 2008. – No. 1 (17). – Pp. 99-101.

14. Lakeyev A.V., Noskov S.I. A description of the set of solutions of a linear equation with interval defined operator and right-hand side // *Doklady Mathematics*. – 1993. – Т. 47. – No. 3. – P. 518.
15. Lakeyev A.V., Noskov S.I. The least module method for linear regression: the number of zero approximation errors // *Modern Technologies. System analysis. Modeling*. – 2012. – No. 2 (34). – Pp. 48-50.
16. Noskov S.I. A generalized criterion for the consistency of behavior in regression analysis // *Information technology and mathematical modeling in the management of complex systems*. – 2018. – No. 1 (1). – Pp. 14-20.
17. Noskov S.I. On the method of mixed estimation of linear regression parameters // *Information technologies and mathematical modeling in the management of complex systems*. – 2019. – No. 1 (2). – Pp. 41-45.
18. Noskov S.I. Identification of parameters of a piecewise linear risk function // *Transport infrastructure of the Siberian region*. – 2017. – Т. 1. – Pp. 417-421.

Информация об авторах

Сергей Иванович Носков – д. т. н., профессор, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: noskov_s@irgups.ru

Владислав Николаевич Городенко - магистр кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: vladgorodenko@gmail.com

Authors

Sergey Ivanovich Noskov, Doctor of Technical Science, Professor, the Subdepartment Information systems and information security, Irkutsk State Transport University, Irkutsk, e-mail: noskov_s@irgups.ru

Vladislav Nikolaevich Gorodenko, master, the Subdepartment Information systems and information security, Irkutsk State Transport University, Irkutsk, e-mail: vladgorodenko@gmail.com

Для цитирования

Носков С.И., Городенко В.Н. Программный комплекс антиробастного оценивания параметров линейного регрессионного уравнения // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2019. – №4(5). – С. 49-55. DOI: 10.26731/2658-3704.2019.4(5).49-55 – Режим доступа: <http://ismm-irgups.ru/toma/45-2019>, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 20.12.2019)

For citations

Noskov S.I., Gorodenko V.N. Programmyj kompleks antirobastnogo ocenivaniya parametrov linejnogo regressiionnogo uravneniya [Software complex of antirobastic evaluation of parameters of a linear regression equation] // *Informacionnyye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal* [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2019. No. 4(5). P. 49-55. DOI: 10.26731/2658-3704.2019.4(5).49-55 [Accessed 25/02/19]